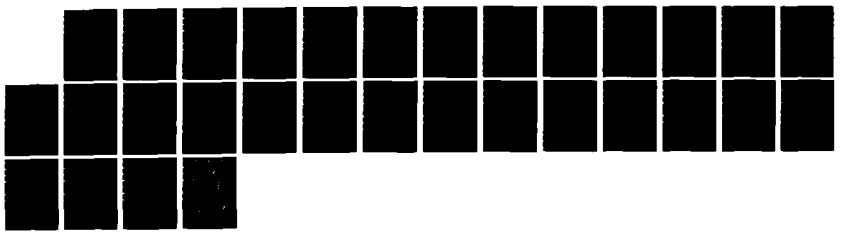
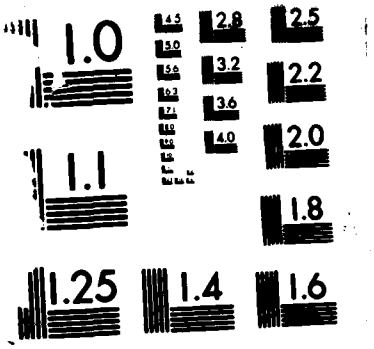
AD-A165 999 DESIGN OF A MASTER LEXICON(U) UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST S CUMMING FEB 86 ISI/RR-85-163 MDA903-81-C-0335 1/1

UNCLASSIFIED F/G 5/7 NL





(12)

ISI Research Report
ISI/RR-85-163
February 1986

University of Southern California

Susanna Cumming

Design of a Master Lexicon

INFORMATION SCIENCES INSTITUTE

213/822-1511
4676 Admiralty Way/Marina del Rey/California 90292-6695

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A165 999

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT This document is approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ISI/RR-85-163	5. MONITORING ORGANIZATION REPORT NUMBER(S) ----------
6a. NAME OF PERFORMING ORGANIZATION USC/Information Sciences Institute 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION ----------
6c. ADDRESS (City, State and ZIP Code) 4676 Admiralty Way Marina del Rey, CA 90292	7b. ADDRESS (City, State and ZIP Code) ----------
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Advanced Research Projects Agency 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903 81 C 0335
8c. ADDRESS (City, State and ZIP Code) 1400 Wilson Boulevard Arlington, VA 22209	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) Design of a Master Lexicon (Unclassified)	
12. PERSONAL AUTHOR(S) Cumming, Susanna	
13a. TYPE OF REPORT Research Report 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 1986, February 15. PAGE COUNT 30
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
09	02	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
artificial intelligence, computational linguistics, dictionary, English, functional grammar, grammar, Janus, LapItUp, lexical acquisition,
(Continued on reverse)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This document describes the design for a "Master Lexicon" which is intended to serve all the lexical needs of the two grammars (RUS, an ATN parser, and Nigel, a systemic production grammar) of the JANUS natural language project. Each grammar had previously been using its own lexicon, and these two lexicons were significantly different, both in terms of content and in terms of design. The Master Lexicon incorporates features of each of the previous lexicons, significantly Nigel's feature hierarchy and RUS's treatment of morphology and cross-indexing; plus it includes several novel features, most notably a new lexical acquisition interface which allows the non-specialist user to add lexical items. Nigel is able to access the Master Lexicon directly, while a set of translation rules (given in the appendix) transform ML lexical entries to lexical entries of the style used by RUS.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Victor Brown Sheila Coyazo	22b. TELEPHONE NUMBER (Include Area Code) 213-822-1511 22c. OFFICE SYMBOL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

18. Subject Terms (continued)

lexical features, lexicon, Master Lexicon, natural language, Nigel, Penman, RUS, syntax, Systemic Linguistics, text generation, vocabulary, word acquisition, wordclass hierarchy, wordclasses

ISI Research Report
ISI/RR-85-163
February 1986

Susanna Cumming

Design of a Master Lexicon

INFORMATION SCIENCES INSTITUTE

213/822-1511
4676 Admiralty Way/Marina del Rey/California 90292-6695

This research is supported by the Defense Advanced Research Projects Agency under Contract No. MDA903 81 C 0335. Views and conclusions contained in this report are the author's and should not be interpreted as representing the official opinion or policy of DARPA, the U.S. Government, or any person or agency connected with them.

Table of Contents

List of Tables

1. Introduction

This report concerns the design for a Master Lexicon (ML) which will serve both a systemic production grammar, Nigel, and an ATN parsing grammar, RUS. A lexicon is the repository for all the idiosyncratic information which must be specified about each word in the language. This information may be organized in various ways, and Nigel and RUS, in the lexicons they have formerly used, have in many cases adopted very different ways of organizing this information. Therefore, the primary task in the design of a Master Lexicon is to identify the general issues and problems in lexicon organization, compare how the two lexicons handle these issues and problems, and decide whether to adopt one of the existing solutions (and if so, which) and where to innovate a new solution. Where the latter course is taken, the new solution adopted is one dictated by concerns of generality, consistency, completeness, and linguistic motivation; thus the ML should be more than simply an intermediary between the two grammars.

Both grammars require of a lexicon the specification of four kinds of information, which must be indicated in the entry for each lexical item. The first is syntactic information: each entry must contain information which will specify the kinds of constructions the word may enter into. This information is generally expressed in a list of features associated with each entry. Thus the verb "arise" might have the features Verb and Intransitive, among others. The second type is morphological information, that is, information that has to do with the forms of the word. Thus, the entry for the stem "arise" must contain the information that the third person singular form is "arises", the past tense is "arose", the present participle is "arising", and the past participle is "arisen". The third type of information has to do with collocations, i.e. whether the word should be related to some other word or phrase in the lexicon: thus "decide" is related to the compound verb "decide on". The fourth type of information is semantic: each word must carry an indication of what concepts in the semantic network it is associated with. (This aspect of lexical specification will only be briefly mentioned in this report.)

Since morphological, syntactic, and collocational information is organized in different ways in Nigel and RUS, the next two sections of this report will describe the organization of the lexicons currently being used by the two grammars, showing how each handles these issues; these sections will provide the background for an understanding of the decisions we've made about the design of the Master Lexicon. The fourth section will present the results of those decisions in a comprehensive overview of the structure of the Master Lexicon.

In addition to resolving general issues of lexicon structure, the Master Lexicon must also resolve particular issues of what features are to be used in lexical entries. As mentioned above, the lexicon is in principle the repository for all the idiosyncratic information about each word in the language; in practice, it is the repository for all the idiosyncratic information a particular grammar needs about words in order to accomplish its task. While certain properties of words are likely to be considered both relevant and idiosyncratic by any grammar, nonetheless the kind of information which is relevant, and even what is to

be considered idiosyncratic and what is general, differs depending on the needs and capabilities of a particular grammar. Thus, a major problem facing the design of a Master Lexicon is to determine where the lexical categories referred to by the two grammars are different, and where they are really the same in spite of inevitable differences in terminology. Once a set of features sufficient for all the needs of both grammars has been arrived at, a set of rules is necessary to translate ML categories into the categories currently used by the existing grammars; the fifth section of this report describes those rules.

Finally, an easily accessible interface for adding words to the lexicon with complete and correct feature specifications is needed. This is a fundamental part of a system designed for transportability, since every domain (and to a lesser extent every user) will have different needs. The sixth section describes our design for this interface.

1.1. Word and Sense

Before these questions can be addressed, however, a short discussion of what is meant by a "word" and what is meant by a "word sense" is in order. The most natural conception of a word is the orthographic one of (roughly) a continuous string of characters with a space at each end, and the simplest idea of a dictionary assigns one lexical entry to every word in the language. However, there are two kinds of cases where this idealization breaks down. The first is the case of homonymy: there are many cases in natural language of two words which are spelled the same but have usages sufficiently distinct to require two different feature specifications. "Intimate" is a good example: it can be either a verb or an adjective, with entirely different meanings. The second is the case of collocation: there are many cases where a sequence of orthographic words has grammatical properties which are not derivable from the words taken individually. A few examples are "as well as", "Long Island", and "decide on". In order to be fully explicit, therefore, we need to be able to distinguish an orthographic word, or spelling (a continuous string of characters), from a lexical item (a string of characters which has its own lexical entry)[1]. A single lexical item may comprise more than one orthographic word, and a single orthographic word may correspond to more than one lexical item. In what follows, the distinction between "lexical item" and "spelling" will be maintained.

Even given the idea that a single spelling may correspond to more than one lexical item, we still need to allow for the possibility of a single lexical item which has more than one word sense. This situation will arise when a spelling has two meanings which are not distinguished by any syntactic feature, but are represented by distinct concepts in the semantics. For example, a "mouse" which is a kind of animal and a "mouse" which is part of a computer will not need to be distinguished by any grammatical feature, but they will occupy very different positions in the semantic net. Therefore the term "senses of a word" will

---

[1] The criteria used to determine when a word has sufficiently divergent usages to justify multiple entries are different for Nigel and RUS; see sections 3.3 and 2.3.

be used to refer to a word with multiple semantic pointers, as distinct from "multiple lexical entries for an orthographic word".

2. The Structure of the RUS Lexicon

My information about the RUS lexicon comes primarily from the documentation provided in [Bates 83], [Bates 84a], and [Bates 84b].

2.1. Syntax

The RUS lexicon has a relatively complex structure. The syntactic information is divided into four different types, expressed in the lexicon by "categories", "features", "properties", and "morphological specifications" (morph specs).

Categories correspond roughly to the traditional parts of speech. Unlike features, they are obligatory: every lexical item must have a category. N (noun), V (verb), and ADJ (adjective) are all categories, but there are also less traditional categories like QPRO (question pronoun), PUNCT (punctuation), and SHORTANSWER (interjections etc.)

Features, on the other hand, cooccur more or less freely (subject to certain restrictions), and are not obligatory -- a word doesn't need to have any features. Features mainly express distributional restrictions on words; for example, verbs have features like BITRANSITIVE (for verbs that take two noun phrase objects but don't allow dative movement, like "cost" in "The coffee cost me a dollar"), NPTOCOMP (for verbs that occur with an object followed by an infinitive complement, like "advise" in "I advise you to come"), and THATREQUIRED (for verbs that take a "that"-complement in which the "that" is obligatory, like "declare" in "I declared that you had been here"). The following examples illustrate the use of categories and features in a RUS lexical entry (features are italicized, categories are boldfaced):

[PUTDICTENTRY 'TURF (QUOTE (FEATURES (*NON-COUNT*) **N** -S))]

[PUTDICTENTRY 'THREATEN (QUOTE (FEATURES (*INTRANSTOCOMP PASSIVE TRANS*) **V** S-ED))]

Properties are similar to features, except that a property has a value. CASEPREPS is a property which indicates what prepositions a verb can take and what case roles they mark with respect to the verb; the value of CASEPREPS is a list of pairs of a preposition and a case role. SUBSTITUTE is a property for abbreviations and synonyms which has as its value another entry which is substituted during the parsing process. COMPOUNDS is a property which has as its value a compound entry whose first word is the current entry. The primary use of properties, as these examples show, is to express various relationships between separate entries in the lexicon. The uses of CASEPREPS and COMPOUNDS are illustrated in the following example (the property name is boldfaced, while the property values are italicized):

[PUTDICTENTRY 'TRANSPORT (QUOTE (FEATURES (PASSIVE TRANS)
CASEPREPS *((FROM SOURCE) (TO DESTINATION))* N -S V S-ED))]

[PUTDICTENTRY 'THANKS (QUOTE (**COMPOUNDS** *((TO THANKS\TO))*
N (THANKS (NUMBER PL)))))]

[PUTDICTENTRY 'THANKS\TO (QUOTE (PREP *))]

Morph specs are mostly used to carry information about the inflected forms of a word. Each category specification has a morph spec, either the default morph spec "**" or a more or less complex expression indicating what inflectional type the entry is, what other entries the entry is related to by inflection, or sometimes just idiosyncratic information about the word, as in the morph spec for a number word, which is simply the arabic numeral referred to by the word. In the above examples of RUS dictionary entries, the morph specs immediately follow the category.

2.2. Morphology

Morphological information in the RUS lexicon is carried by the morph specs. Regular inflectional information for open-class items (nouns, verbs, and adjectives) is indicated by a feature on the stem entry specifying the suffix spelling. For noun stems, for instance, there are two possible regular specs which indicate the plural form, -ES and -S. NOINFLECTIONS is used for words that aren't inflected, like "linguistics" or "doctoral". Words which are inflected irregularly are marked by the spec IRR[2]; the inflected form has its own entry, which has a spec indicating what the stem is and what form of the stem it is. Irregularly inflected noun entries have a NUMBER specification, and irregularly inflected verb entries have a PNCODE (agreement) specification and a TNS (tense) specification. In the following examples, the morph specs are in italics.

[PUTDICTENTRY 'TOOTH (QUOTE (N *IRR*))]

[PUTDICTENTRY 'TEETH (QUOTE (N *(TOOTH (NUMBER PL))*))]

[PUTDICTENTRY 'TAKE (QUOTE (FEATURES (INDOBJ INTRANS PASSIVE TRANS)
V *(TAKE (PNCODE X3SG) (TNS PRESENT) (UNTENSED))*))]

[PUTDICTENTRY 'TAKEN (QUOTE (V *(TAKE (PASTPART))*))]

[PUTDICTENTRY 'TAKING (QUOTE (V *(TAKE (PRESPART))*))]

[PUTDICTENTRY 'TOOK (QUOTE (V *(TAKE (TNS PAST))*))]

[PUTDICTENTRY 'UNMARRIED (QUOTE (ADJ *NOINFLECTIONS*))]

Paradigms are supplied in this way for nouns (singular and plural forms), adjectives (absolutive[3],

[2]Even in cases where only one inflected form of a verb is irregular, there must be separate entries for all of its forms.

[3]This form is often referred to in traditional grammar as the "positive" form of an adjective.

comparative and superlative forms), and verbs (stem, third person singular, simple past, past participle and present participle forms).

Some closed classes have morph specs as well: determiners and pronouns have a NUMBER specification, and pronouns also are specified for case. For the details of RUS morphological specifications, see the translation rules in section 10 of the appendix.[4]

2.3. Multiple Word Entries

As mentioned above, the primary use of properties is to provide a pointer or cross-reference to another lexical entry. This is necessary when an orthographic word can be part of a compound phrase of some sort which has a meaning and syntax distinct from that of the word when it occurs alone. For example, "PARTICLES" is a verb property which may have one or more values depending on the verb-particle constructions possible with the verb. Thus, the verb "hand" has the property PARTICLES with the values "in", "out" and "over", which serve as pointers to the separate lexical entries for "hand in", "hand out" and "hand over". Other properties which can be used similarly as pointers are IMMOVABLEPARTICLES (like PARTICLES, only they don't permit particle movement), COMPOUNDS (any string of orthographic words that has an idiosyncratic meaning when taken together, and each of which has its own lexical entry) and MULTIPLES (like compounds except that only the first word must have its own lexical entry).

2.4. Homonyms

Another property, SUBSTITUTE, can be used to handle homonymy. Sometimes an orthographic word has two distinct meanings which are spelled the same way. In these cases, it is often undesirable to assign all the same features to both meanings, and some method of distinguishing the two uses of the word is required. An example of this is "will", which has two verbal meanings; one of them has regular inflections, "will wills willed willed willing", while the other has the irregular (and defective) pattern typical of modals, "will will would". One solution is to have two separate entries, e.g. "will" and "will-modal"; "will" has among its specifications the property SUBSTITUTE with the value "will-modal", which serves to cross-index the two entries. This property is not used very widely in RUS for this purpose, though;[5] generally a homonym simply has in its feature list all of the features associated with each sense mixed together. Thus "control" has the categories N and V, and the features NONCOUNT (a noun feature), and TRANS and PASSIVE (verb features).

---

[4]Much of my information about RUS morphology specification comes from the draft document [Ingria 85].

[5]Substitute is principally used to replace abbreviations with their full form; for example, "U.S." has the property SUBSTITUTE with the value "united-states".

3. The Structure of the Nigel Lexicon

The Nigel lexicon is much simpler than the RUS lexicon, in that all syntactic and morphological information is encoded by a uniform device, the feature. However, in another respect the Nigel lexicon is more complex. The Nigel features are arranged in a "wordclass hierarchy", such that every feature is associated with one or more wordclasses which have superclass and subclass relationships with other classes in the hierarchy. This hierarchy expresses relationships among features, such as the fact that some nouns are common nouns, and some common nouns are plural. In the original Nigel system, every lexical item was associated with a single wordclass, and the feature specification of that item was obtained by finding each of the features associated with each superclass of the class the item was in. One result of this structure is that Nigel assumes very rigid limits on which features can cooccur on the same word; only the features which can occur on some path from the root to a leaf of the tree can ever be assigned to the same word.

3.1. Syntax

Because of the functional nature of Nigel's grammar, the kinds of categories referred to by the grammar tend to be semantically-characterized classes which share clusters of syntactic patterns; thus verbs have features such as Reaction ("like", "grieve", "please"), Cognition ("amaze", "remind", "understand"), and Perception ("hear", "notice", "strike"). The complementation possibilities of each verb are taken to be predictable from the semantic class of the verb. Thus the wordclass hierarchy used by Nigel has some similarity to a semantic taxonomy.

3.2. Morphology

Nigel originally contained a separate entry for each morphological form of a lexical item; there was no separate mechanism for this kind of information, morphological distinctions being indicated by features which were exactly like syntactic features (thus "accompanies" had, among others, the features Thirdperson, Singular, Stateverb, and Verb). Since the development of the Master Lexicon, inflectional rules have been added to Nigel which use the sort of inflectional information found in the RUS dictionary to inflect stems.

3.3. Multiple Word Entries

Some of the combinations of the sort handled by the RUS "properties" IMMOVABLEPARTICLES, COMPOUNDS, and MULTIPLES are handled by a distinction in the Nigel lexicon between "word names" and "spellings". Every word in the lexicon has both a word name and a spelling. In many cases the two will be the same; they differ when a distinction must be made between lexical item and orthographic word. In the case of multiple word entries, the "word name" may consist of more than one word treated as a unit, while the spelling corresponds to a list of orthographic words. For example, "act as" has the word name "actas" and the spelling "act as". However, unlike the RUS system, the Nigel

system provides no cross-reference between the entries for "act as" and "act".

3.4. Homonyms

Homonyms are also handled by the wordname/spelling distinction: two words may have the same spelling but different wordnames, so that "besides" the subordinator has the wordname "besides", while "besides" the adverb has the wordname "besidesadv". Thus, the features of the adverb and the subordinator senses of "besides" are kept distinct.

4. The Structure of the Master Lexicon

The Master Lexicon combines some aspects of Nigel's organization with some aspects of RUS's, and has some entirely new characteristics. What follows is a sketch of how particular kinds of information are represented in the ML; for a complete description of every open class feature and property, and the complete ML wordclass hierarchy, see [Cumming 86].

4.1. Syntax

The arrangement of the features of the ML is based on Nigel's system, with extra features added to cover the distinctions made in RUS but not in Nigel. This is primarily because the structured nature of the Nigel system provides more information than the relatively unstructured RUS system. For the sake of flexibility, the Nigel practice of not distinguishing categories from features has been retained. The ML feature system thus contains all of the existing Nigel features with the same names and in the same relationship to each other as in Nigel, but also contains new features motivated by RUS. Some of these new features have the same names as they do in RUS, but many have been renamed for the sake of clarity and terminological consistency.

The ML is different from the Nigel lexicon in that about half of the features cooccur freely with the others; this means that a word may be in more than one wordclass in the ML. Information about restrictions on cooccurrence is contained in the wordclass hierarchy by means of the "group" convention: if several wordclasses are in a group, it means that a given lexical item can't belong to more than one wordclass in the group. For example, "propernoun" and "commonnoun" are in a group; that means that a noun cannot be both proper and common. There are two different kinds of groups, called "Group1" and "Group0". If two or more features are in a Group1 relationship, a lexical item must have exactly one of the features. If two or more features are in a Group0 relationship, a lexical item may have either none of the features or one of them. "propernoun" and "commonnoun" are in a Group1: this means that every noun must be specified as either proper or common.

The ML is also different from Nigel in that, like RUS, it contains properties. Since properties, like features, have taxonomic dependencies (e.g. only verbs can have PARTICLES), each property has a

corresponding feature with the same name which has a place in the wordclass hierarchy; however, each property is also associated with a value, which differs depending on the item that has that property.

Unlike both Nigel and RUS, the ML has explicit negative feature specifications; features which could have been assigned to a lexical item without violating wordclass structure constraints, but were not, receive negative values in the feature specification. These are feature names prefaced by NOT- (or NONE-OF-, in the case of Group0s where none of the features were chosen). The presence of these negative feature values enables the ML consistency checker to confirm the completeness of a given lexical entry with respect to a given wordclass hierarchy.

4.2. Morphology

Morphology is handled in the ML much as it is in RUS, with two exceptions: 1) what is handled in RUS by morph specs is handled in the ML by a combination of properties and features, and 2) rather than receiving their own entries as they do in RUS, irregularly inflected forms are simply contained as property values in the stem entry for lexical item. Translation rules are used to create lexical entries for these irregular forms for RUS.

4.2.1. Open Class Inflection

There are four ways of specifying affixation for verbs, nouns, and adjectives:

1. A stem that doesn't take any inflections at all is assigned the feature Noinflections. This applies to nouns that have no plural forms, adjectives that don't enter into comparative constructions, and to verb-particle combinations, to keep from getting entries like "look ups".

2. A stem that has one of several regular affixation patterns is assigned a feature to show what pattern it takes. The only stem spelling alternations which are handled as regular are e~0 (decide/deciding), y~i (reply/replies), and final consonant doubling (signified by an asterisk in the feature names: run/running, flat/flattest); other words are considered regular only if the spelling of the stem is invariant. The regular features that are recognized are S and Es for nouns; S-d, Es-ed, S-ed, and S-*ed for verbs; and R-st, Er-est and *Er-*est, and More-most for adjectives.

3. A stem with irregular forms is assigned the feature "Irr". Its irregular inflected forms receive their own entries, which are given a feature to indicate which form they are and cross-indexed to the stem by the "stem" property, which has as its value the name of the stem. In addition, the stem carries properties which specify what its inflected forms are.

4. Some verbs are partially regular, i.e. they have regular third person singular and present participle forms but irregular past participle and past forms. These will be given one of the features S-irr, Es-irr, or *-Irr depending on which third singular form they take. Partial irregularity is not allowed for in RUS; there, if one part is irregular, each part must be specified. Translation rules take the partially regular features to RUS features (see section 11.2 of the appendix).

The use of morphological features and properties is illustrated in table 4-1.

```
 :name 'SPELL
 :spelling "spell"
 :features '(NOUN NOT-NOMINALIZATION COMMON COUNTABLE S NONSUBSTITUTE
NOT-THATCOMP)
 :properties '()

 :name 'ADDENDUM
 :spelling "addendum"
 :features '(NOUN NOT-NOMINALIZATION COMMON COUNTABLE NONSUBSTITUTE
NOT-THATCOMP IRR PLURALFORM)
 :properties '((PLURALFORM "addenda" ))

 :name 'CHARGE
 :spelling "charge"
 :features '(VERB INFLECTABLE UNITARYSPELLING S-D LEXICAL
NOT-CASEPREPOSITIONS NOT-THATCOMP NOT-PARTICIPLECOMP OBJECTPERMITTED
NOT-PASSIVE EFFECTIVE DOVERB DISPOSAL BITRANSITIVE NOT-OBJECTNOTREQUIRED
NOT-SUBJECTCOMP NOT-QUESTIONCOMP NOT-TOCOMP NOT-MAKECOMP NOT-COPULA
NOT-ADJECTIVECOMP NOT-BAREINFINITIVECOMP)
 :properties '()

 :name 'STAND
 :spelling "stand"
 :features '(VERB INFLECTABLE UNITARYSPELLING LEXICAL
NOT-CASEPREPOSITIONS S-IRR PASTFORM EDPARTICIPLEFORM NOT-OBJECTPERMITTED
MIDDLE DOVERB BEHAVIOUR NONE-OF-BITRANSITIVE-INDIRECTOBJECT
OBJECTNOTREQUIRED OBJECTNOTPERMITTED NOT-SUBJECTCOMP NOT-THATCOMP
NOT-PARTICIPLECOMP NOT-QUESTIONCOMP NOT-TOCOMP NOT-MAKECOMP
NOT-ADJECTIVECOMP NOT-BAREINFINITIVECOMP)
 :properties '((PASTFORM "stood" )(EDPARTICIPLEFORM "stood" ))

 :name 'SIMPLE
 :spelling "simple"
 :features '(ADJECTIVE NOT-CASEPREPOSITIONS R-ST DEGREE
COMPLEMENTPERMITTED TOCOMP FORNPPERMITTED SUBJECTHOLD NOT-SUBJECTCOMP
NONE-OF-APPROPRIATENESS-POSSIBILITYPROPERTY-OBVIOUSNESS NOT-THATCOMP
NOT-PREDICATEONLY)
 :properties '()

 :name 'GOOD
 :spelling "good"
 :features '(ADJECTIVE CASEPREPOSITIONS IRR DEGREE COMPLEMENTPERMITTED
NOT-PREDICATEONLY FORNPPERMITTED SUBJECTHOLD SUBJECTCOMP THATCOMP
TOCOMP NONE-OF-APPROPRIATENESS-POSSIBILITYPROPERTY-OBVIOUSNESS
SUPERLATIVEFORM COMPARATIVEFORM)
 :properties '((SUPERLATIVEFORM "best" )(COMPARATIVEFORM "better" ))
```

Table 4-1: Use of Morphological Features

Besides the inflectional relationships exemplified above, Nigel recognizes an additional relationship between open class entries: the relationship between a verb and its nominalization. This too is handled as a "property" in the ML: a noun which is a nominalization will have it listed in the noun entry and cross-

indexed with a separate verb entry, as in table 4-2. The noun entry will have in addition whichever of the noun features are appropriate[6].

```
:name 'RESPONSE
:spelling "response"
:features '(NOUN COMMON COUNTABLE NONSUBSTITUTE S THATCOMP
NOMINALIZATION)
:properties '((NOMINALIZATION RESPOND ))

:name 'RESPOND
:spelling "respond"
:features '(VERB INFLECTABLE UNITARYSPELLING S-ED LEXICAL
CASEPREPOSITIONS MIDDLE SYMBOLICVERB SPEAKING INDIRECTOBJECT
OBJECTNOTREQUIRED NOT-OBJECTNOTPERMITTED NOT-TOCOMP NOT-QUESTIONCOMP
NOT-PARTICIPLECOMP THATCOMP THATREQUIRED NOT-SUBJUNCTIVEREQUIRED
NOT-SUBJECTCOMP NOT-OBJECTPERMITTED NOT-MAKECOMP NOT-ADJECTIVECOMP
NOT-BAREINFINITIVECOMP)
:properties '((CASEPREPOSITIONS TO WITH BY ))
```

Table 4-2: Use of Nominalization

4.2.2. Closed Class Inflection

RUS requires the specification of number, stem and case on some kinds of determiners and pronouns[7].

Interrogative and deictic determiners require the specification of number, with possible values singular, plural, countable, and uncountable (or any combination, except countable and uncountable are mutually exclusive). See table 4-3.

Possessive determiners (i.e. "my, your" etc.) require the specification of the (subject) form of their stem, as shown in table 4-4.

Other pronouns require the specification of the (subject) form of their stem[8], their number (possible values singular, plural, or singular/plural), and their case (possible values subject, object, or both), as shown in table 4-5.

One other property required to get RUS morph specs right (although it isn't really a morphological property) is "Arabicprop", used for giving the Arabic numeral associated with an integer or an ordinal.

---

[6]Currently the only type of nominalization handled by Nigel is the "process" type; the same mechanism can easily be extended to handle other types as the grammar is modified to produce them.

[7]In the current version of the RUS dictionary, these specs are distributed rather erratically. The principles enumerated below, and the translation rules I've written, reflect my understanding of how things ought to be, not how they currently are in many cases. There is probably some redundancy in the specifications here, which could be eliminated by complicating the translation rules; for example, all pronouns which can be subjects have themselves as the value of the Stem property; all pronouns which are Possessive have exactly the same values for case, number, and stem. However, I am reluctant to propose these complications until the RUS requirements become clearer.

[8]Pronouns which can be subjects and objects have themselves as the value of the Stemform property.

```
:name 'WHICH
:spelling "which"
:features  '(DETERMINER NOT-POSSESSIVEDETERMINER INTERROGATIVE NUMBER)
:properties  '((NUMBER SINGULAR PLURAL UNCOUNTABLE ))

:name 'HOW-MUCH
:spelling "how much"
:features  '(DETERMINER NOT-POSSESSIVEDETERMINER INTERROGATIVE NUMBER)
:properties  '((NUMBER UNCOUNTABLE ))

:name 'THE
:spelling "the"
:features  '(DETERMINER NOT-POSSESSIVEDETERMINER DEICTIC NUMBER)
:properties  '((NUMBER SINGULAR PLURAL UNCOUNTABLE ))

:name 'EVERY
:spelling "every"
:features  '(DETERMINER NOT-POSSESSIVEDETERMINER DEICTIC NUMBER)
:properties  '((NUMBER SINGULAR COUNTABLE ))
```

Table 4-3: Specification of Interrogative and Deictic Determiners

```
:name 'MY
:spelling "my"
:comment  ""
:features  '(DETERMINER NOT-NUMBER DEICTIC POSSESSIVEDETERMINER STEMFORM)
:properties  '((STEMFORM I ))
```

Table 4-4: Specification of Possessive Determiners

```
:name 'HE
:spelling "he"
:features '(PRONOUN CASE STEMFORM NUMBER
NONE-OF-INTERROGATIVE-NOPOSTMODIFIERS-INDEFINITEPRONOUN-LOCATION-
SUGGESTIVEPARTICLE-POSSESSIVEPRONOUN)
:properties  '((CASE SUBJECT )(STEMFORM HE )(NUMBER SINGULAR ))

:name 'HIM
:spelling "him"
:features  '(PRONOUN CASE STEMFORM NUMBER NOPOSTMODIFIERS)
:properties  '((CASE OBJECT )(STEMFORM HE )(NUMBER SINGULAR ))

:name 'THEIRS
:spelling "theirs"
:features  '(PRONOUN CASE STEMFORM POSSESSIVEPRONOUN NUMBER)
:properties  '((CASE SUBJECT OBJECT )(STEMFORM THEY )(NUMBER PLURAL ))

:name 'ANYONE
:spelling "anyone"
:features  '(PRONOUN INDEFINITEPRONOUN CASE STEMFORM NUMBER)
:properties  '((CASE SUBJECT OBJECT )(STEMFORM ANYONE )(NUMBER SINGULAR ))
```

Table 4-5: Specification of Pronouns

```
:name 'FOURTEEN
:spelling "fourteen"
:features '(DETERMINER NOT-NUMBER NOT-POSSESSIVEDETERMINER NUMERATIVE
COUNTABLE CARDINAL NUMERAL ARABICPROP)
:properties '((ARABICPROP 14 ))

:name 'FOURTEENTH
:spelling "fourteenth"
:features '(ORDINAL ARABICPROP)
:properties '((ARABICPROP 14 ))
```

Table 4-6: Specification of Numbers

4.3. Multiple Word Entries

Since cross-indexing of a compound with its first member (as is required in RUS) is redundant in the context of the Master Lexicon, the RUS properties COMPOUNDS and MULTIPLES will be added to the appropriate entries by the translation rules. The distinction between the properties PARTICLES and IMMOVABLEPARTICLES, however, is not redundant; therefore, the entry for a verb-particle compound contains a feature indicating whether the particle is movable or not, although cross-indexing is still handled by translation rule. (The ML consistency maintenance checker requires that there be a lexical entry for the verb portion of a verb-particle combination; this guarantees that RUS will be able to create the cross-index.) CASEPREPS present a different problem, since with case prepositions the verb-preposition combination does not have its own lexical entry. For this reason case prepositions will be handled as they are in RUS, i.e. as a property, except that a caserole name will not be supplied[9]. Certain parts of what RUS expects as the values for these properties are redundant; these parts will be supplied by rule, as specified in section 11.2 of the appendix. The use of PARTICLES and CASEPREPOSITIONS is illustrated in table 4-7.

4.4. Homonyms

The ML approach to homonymy reflects the Nigel one: every different sense of a word will have a separate entry. This is a desirable feature, since it is not only more linguistically motivated, but incorrect feature specifications can result from failing to distinguish senses, due to the fact that same features can occur in more than one place in the wordclass hierarchy. For example, the word "intimate" can be either a verb or an adjective. As a verb, it should have the feature "Thatcomp" and the property "(Casepreps TO)". As an adjective, it shouldn't have the feature "Thatcomp" (although this is a possible adjective feature), and it should have the property "(Casepreps WITH)". In the ML, there would be two different lexical entries, e.g. "INTIMATE" and "INTIMATE-ADJ". The RUS translation rules will operate in such a way that both of these entries are merged into a single entry for INTIMATE.

---

[9] As mentioned in 2.1, at present RUS has as the value of CASEPREPS a pair of a preposition and a caserole name; however, the caserole name is not currently used anywhere in the grammar, and cannot be supplied appropriately except in the context of a particular knowledge base.

```
 :name 'LEAD
 :spelling "lead"
 :features '(VERB INFLECTABLE LEXICAL NOT-CASEPREPOSITIONS
OBJECTPERMITTED NOT-TOCOMP NOT-QUESTIONCOMP NOT-PARTICIPLECOMP
NOT-MAKECOMP NOT-BAREINFINITIVECOMP NOT-COPULA PASSIVE NOT-THATCOMP
NOT-ADJECTIVECOMP NONE-OF-BITRANSITIVE-INDIRECTOBJECT DOVERB DISPOSAL
EFFECTIVE OBJECTNOTREQUIRED NOT-OBJECTNOTPERMITTED SUBJECTCOMP
UNITARYSPELLING S-IRR PASTFORM EDPARTICIPLEFORM)
 :properties '((PASTFORM "led" )(EDPARTICIPLEFORM "led" ))

 :name 'LEAD-TO
 :spelling "lead to"
 :features '(VERB INFLECTABLE LEXICAL NOT-CASEPREPOSITIONS
OBJECTPERMITTED TOCOMP NOT-SUBJECTLOWERING FORNPPERMITTED
NOT-NOLOWERINGVERB NOT-QUESTIONCOMP NOT-PARTICIPLECOMP NOT-MAKECOMP
NOT-BAREINFINITIVECOMP NOT-COPULA PASSIVE NOT-THATCOMP NOT-ADJECTIVECOMP
INDIRECTOBJECT RELATIONAL CIRCUMSTANTION CAUSE EFFECTIVE
NOT-OBJECTNOTREQUIRED SUBJECTCOMP COMPOUNDSPELLING PARTICLES
NOT-NP-TOCOMPVERB)
 :properties '()

 :name 'BRING
 :spelling "bring"
 :features '(VERB INFLECTABLE LEXICAL CASEPREPOSITIONS OBJECTPERMITTED
NOT-TOCOMP NOT-QUESTIONCOMP NOT-PARTICIPLECOMP NOT-MAKECOMP
NOT-BAREINFINITIVECOMP NOT-COPULA PASSIVE NOT-THATCOMP NOT-ADJECTIVECOMP
INDIRECTOBJECT DOVERB DISPOSAL EFFECTIVE NOT-OBJECTNOTREQUIRED
NOT-SUBJECTCOMP UNITARYSPELLING S-IRR PASTFORM EDPARTICIPLEFORM)
 :properties '((PASTFORM "brought" )(EDPARTICIPLEFORM "brought")
(CASEPREPOSITIONS WITH TO ))
```

Table 4-7: Handling of Multiple Word Entries

5. Accommodation Rules

5.1. Syntactic Features: Rules for RUS

The differences between RUS's method of specifying syntactic information and that adopted in the ML require a set of derivation rules which will take an ML feature specification for a word and translate it into a RUS specification. These rules are given in full in Appendix 11.2.

As explained above, a RUS lexical entry consists of the following fields: 1) the lexical item head, a unique word; 2) one or more category names, each optionally followed by a "spec"; 3) FEATURES, followed by a list of feature names; 4) one or more property names, followed by a list of values for each property. Only the first two fields are obligatory.[10]

---

[10] Under certain circumstances -- when a word is only present to provide a cross-index to another word via the properties SUBSTITUTE, MULTIPLES, and COMPOUNDS -- the category may be omitted.

An ML lexical entry, on the other hand, consists of the following fields: 1) the word name; 2) the word spelling; 3) a list of features; 4) a list of properties with their values; 5) a pointer to JANUS' semantic network (i.e. a NIKL concept name); 6) some record-keeping information ("sample sentence", comment, editor's name, edit date; these fields have been omitted from the above examples). Fields 1-3 and the example sentence are obligatory. Table 5-1 illustrates a fully specified lexical item:

```
(make-lexical-item
  :name 'REQUEST-NOUN
  :spelling "request"
  :sample-sentence "The window shows a request by Jones"
  :comment "used in example 20"
  :features '(NOUN NOMINALIZATION COMMON COUNTABLE NONSUBSTITUTE S THATCOMP)
  :properties '((NOMINALIZATION REQUEST ))
  :date-of-edit "Thursday the seventh of November, 1985; 6:14:03 pm"
  :editors-name "CUMMING"
  :semantics '(NATURALLANGUAGEREQUESTACTION))
```

Table 5-1: A fully specified ML lexical item

The RUS lexical item head will be derived from the ML spelling, not from the wordname. RUS categories and features will be derived from ML features. RUS morph specs will be derived from ML properties, features, and spellings. RUS properties and property values will be derived from ML properties and property values, with the exception of redundant (cross-indexing) properties which will be computed from combinations of lexical entries.

Each rule has on its right-hand side a ML feature, property, or complex of features and/or properties related by the operators "and" and "or". On its left-hand side it has RUS categories, features, and properties. Some rules (the ones for properties and morph specs) use variables which correspond to the head of a lexical entry or to property values. The feature hierarchy and the rules are structured in such a way that every lexical item will receive a RUS head, category and morph spec and possibly one or more RUS features or properties.

The translation rules given here only support those aspects of RUS lexical specifications which we believe to be functional in the grammar at the present time. Amendments to RUS will require amendments to the rules.

6. Acquisition of Lexical Items

An important part of any computational lexicon is a system for adding words. The ML has a system which allows a user who has no specialized linguistic training to add open-class words to the lexicon. This system is fully described in [Cumming 86]. It has the following features:

It is usable by anyone who has minimal linguistic knowledge (familiarity with terms like "noun", "verb",

"adjective", "subject" etc. should be sufficient) and no familiarity with Nigel and RUS. This goal precludes asking the user directly which features should be assigned to a new lexical item, since considerable familiarity with both general linguistics and the details of Nigel and RUS are required to answer this kind of question appropriately. Instead the user is presented with a series of feature menus; each feature is accompanied by an explanation of what the feature means in the form of tests for whether a word should receive that feature and several examples.

A minimum number of questions will be asked about each word. Since there is a hierarchical dependency relationship among the features, it is not necessary to ask about each feature seperately; the questions which are asked depend on the answers which have been given. Feature grouping also helps reduce the number of steps required; where several features form a group, rather than have a separate yes/no question about each feature, the user can simply select one of the features from a menu. Finally, closed classes (like Month, Possessive (the suffixes "'s" and "'"), Auxiliary and Modal), to which we do not anticipate that a user will want to add, don't need to be queried. Using these principles, the largest number of questions which it should be necessary to ask about any word is under twenty. Verbs which take a variety of complement types are eligible for the most features; for nouns the number is much smaller, around seven. If it were necessary to ask about each feature in the lexicon, more than 230 questions would have to be answered for each word, most of them inappropriate.

In acquiring the morphological forms of a word, the user is presented with the system's best guess at the paradigm, based on the stem (provided by the user) and a simple set of spelling rules. The user can then indicate any forms which are incorrect and substitute the correct forms. The program then takes care of determining the correct morphological features and properties.

Error correction during lexical entry is very simple. The program continuously displays a list of the features that have been chosen; the user may reenter the question tree at any point by mousing a feature in this list. All features dependent on that feature are erased from the list, and the menu associated with that feature is redisplayed.

This general review facility also allows two additional possibilities. One is "reacquisition", the modification of a previously entered lexical item. By giving the name of an existing lexical item, the user has the option of reviewing and modifying its feature list. Alternatively, a user who wishes to enter a new word which has a very similar specification to an existing entry can invoke the definition for the existing entry and modify the feature list in whatever way is appropriate to the new entry.

The user is required to give a "sample sentence" illustrating the intended use of every lexical item entered. This a useful way of helping a user distinguish between multiple senses of a spelling, both while selecting features for a given sense during acquisition and when reviewing a lexical entry. The acquisition program also adds the name of the person who last accessed a lexical item and the date; these fields are

also useful for record-keeping and review purposes.

7. Summary

The Master Lexicon combines features of the RUS lexicon and of the Nigel lexicon. Like Nigel, it makes no distinction between categories and features, and it organizes features in a hierarchy. Like RUS, it contains special kinds of features called "properties", and also morphological features. It is different from both in having a way of including nominalizations in verb entries, and in having a new lexical acquisition interface which enables an untrained user to add items to the lexicon.

I. ML to RUS Translation Rules

The following sections are arranged by the RUS field that is filled by the rule output.

8. WORD

RUS doesn't systematically distinguish separate senses of a word; all senses of a given spelling have a single entry, and there's no equivalent of the unique wordname. Thus the translation rules need to be written to combine all entries with a given word spelling into a single entry; this should be done after the translation rules apply (to prevent rules which take a feature complex as input from applying inappropriately.)

The RUS item should be the ML word spelling, upper-cased, with spaces replaced by \.

9. CATEGORIES

ML FEATURE SPECIFICATION	RUS CATEGORY
Adjectives	
Adjective	ADJ
Verbs	
Verb	V
Pronouns	
(Nopostmodifiers or Substitute)	PRO
(Pronoun and Interrogative)	QPRO
(Preposition or (Pronoun and Location))	PREP
Nouns	
Proper	NPR
Month	MONTH
(Nonsubstitute or Uncountable)	N
Adverbs	
(Adverb and Interrogative)	QWORD
Negative	NEG
Comparativeadverb	COMP
Intensifier	INTENSIFIERADV
(Attitudinalaverb or Manneradverb or Phorictimeadverb or Otheradverb)	ADV
Determiners	
Possessivedeterminer	POSSPRO
(Determiner and Interrogative)	QDET
Deictic	DET
(Nonnumeral or Noncardinal or Uncountable)	QUANTADJ
Linker	

Punctuation	PUNCT
Subordinator	BINDER
Sentenceconjunction	SENTCONJ
Conjunct	CONJ
Other things	
Interjection	SHORTANSWER
Ordinal	ORD
Genitives	POSS
Special	SPECIAL

10. MORPHOLOGICAL SPECIFICATION

The categories which get morph specs are the inflected categories **Noun**, **Verb** and **Adjective**, and the closed class categories **Deictictdeterminer**, **Interrogativedeterminer**, **Possessivepronoun**, **Pronoun**, **Ordinal**, and **Numeral**. Any category which doesn't get a morph spec by these rules should get the default morph spec *. W stands for the wordname of the entry.

10.1. Open Class Specs

Noun, verb, and adjective specs on roots indicate the appropriate endings; specs on irregular inflected forms indicate the stem. "Have" and "be", the only fully irregular verbs, require their own set of rules. Irregularly inflected words require the construction of new lexical entries for RUS; these new entries are enclosed in curly brackets in the following rules. (The translations of the forms of the various lexical items with the spellings "have" and "be" are given in their entirety, since they require unique treatment in both lexicons.)

ML FEATURES AND PROPERTIES	RUS MORPH SPEC
Noinflections	NOINFLECTIONS
Nouns	
S	-S
Es	-ES
Irr; Y; (Pluralform X)	IRR {'W (N (Y (NUMBER PL)))))}
Adjectives	
More-most	*
R-st	R-ST
Er-est	ER-EST
*er-*est	*ER-*EST
Irr; (Comparativeform C); (Superlativeform S)	IRR {'W (ADJ (Y (COMPARATIVE)))))} {'W (ADJ (Y (SUPERLATIVE)))))}
Verbs	
S-d	S-D

Es-ed	ES-ED
S-ed	S-ED
S-*ed	S-*ED
S-irr; (Pastform X); (Edparticipleform Y)	(W (PNCODE X3SG) (TNS PRESENT) (UNTENSED))
Es-irr; (Pastform X); (Edparticipleform Y)	(W (PNCODE X3SG) (TNS PRESENT) (UNTENSED))
	{W (V (X (PNCODE ANY) (TNS PAST)))} {W (V (X (PASTPART)))} {W (V (X (PRESPART)))}

10.2. have

Irr; (Thirdsingularform **has**); (Pastform **had**); (Edparticipleform **had**); (Ingparticipleform **having**)	(HAVE (PNCODE X3SG) (TNS PRESENT) (UNTENSED))
has	{'HAS (V (HAVE (PNCODE 3SG) (TNS PRESENT)))}
had	{'HAD (V (HAVE (PASTPART) (TNS PAST)))}
having	{'HAVING (V (HAVE (PRESPART)))}

10.3. be

Irr; (Firstsingularform **am**); (Secondsingularform **are**); (Thirdsingularform **is**); (Pluralform are); (Firstsingularpastform **was**); (Secondsingularpastform **were**); (Thirdsingularpastform **was**); (Pluralformpastform **were**); (Edparticipleform **been**); (Ingparticipleform **being**)	(BE (UNTENSED))
am Present; Singular; Firstperson; (Stemform **be**)	{'AM (V (BE (TNS PRESENT) (PNCODE 1SG)))}
are Present; Singular; Secondperson; (Stemform **be**)	{'ARE (V (BE (TNS PRESENT) (PNCODE X13SG)))}
Present; Plural; (Stemform **be**)	{'ARE (V (BE (TNS PRESENT) (PNCODE X13SG)))}

is	
Present; Thirdperson; Singular; (Stemform **be**)	{'IS (V (BE (TNS PRESENT) (PNCODE 3SG))))}
was	
Past; Singular; Firstperson; (Stemform **be**)	{'WAS (V (BE (TNS PAST) (PNCODE 13SG))))}
Past; Singular; Thirdperson; (Stemform **be**)	{'WAS (V (BE (TNS PAST) (PNCODE 13SG))))}
were	
Past; Singular; Secondperson; (Stemform **be**)	{'WERE (V (BE (TNS PAST) (PNCODE X13SG))))}
Past; Plural; (Stemform **be**)	{'WERE (V (BE (TNS PAST) (PNCODE X13SG))))}
been	
Edparticiple; (Stemform **be**)	{'BEEN (V (BE (PASTPART))))}
being	
Ingparticiple; (Stemform **be**)	{'BEEN (V (BE (PRESPART))))}

10.4. Closed class specs

Determiner specs indicate number; pronoun specs indcate number, case, and the stem. Pronouns which have the feature Possessive (these are **mine, yours** etc., not **my, your** etc.) have POSS added to their morph spec; this is indicated in the pronoun rule by angled brackets. Ordinal numbers and integers have as their morph spec the corresponding numeral.

It seems reasonable to rename the values for case and number in the ML for the sake of consistency, and translate them into the RUS values.

Interrogative and Deictic Determiners:	
(Number value1 ... valuen)	(NUMBER value1 ... valuen)
Possessive Determiners:	
Possessivedeterminer; (Stemform value)	(value (POSS))
Pronouns	
Pronoun; <Possessive>; (Stemform S); (Number N); (Case C1...Cn)	(S (C1)...(Cn) (N) <(POSS)>)
Numbers:	
Ordinal; (Arabic N)	N
Numeral; (Arabic N)	(N)
Values for Case and Number:	
subject	SUBJ
object	OBJ

countable	NOMASS
uncountable	MASS
singular (for pronouns, not dets)	SG
plural (for pronouns)	PL
singular/plural	SG/PL

11. FEATURES

ML FEATURE SPECIFICATION	RUS FEATURE
Adjectives	
Predicateonly	PREDADJ
Subjectcomp	SUBJCOMP
((Adjective or Verb) and Thatcomp)	THATCOMP
Tocomp	TOCOMP
Subjecthold	SUBJHOLD
Subjectlowering	SUBJLOW
Verbs	
Copula	COPULA
(Perception and Middle)	PERCEIVECOMP
Objectnotrequired	INTRANS
(Verb and Objectpermitted)	TRANS
Passive	PASSIVE
Bitransitive	BITRANSITIVE
Indirectobject	INDOBJ
Seeming	ITSUBJ-THATCOMP
Fornppermitted	FORTOCOMP
(Fornppermitted and Nolowering)	NOLOW
Adjectivecomp	ADJCOMP
Questioncomp	IDQOBJ
Thatrequired	THATREQUIRED
Subjunctiverequired	SUBJUNCTIVEREQUIRED
Np-thatcomp	INDOBJ&THATCOMP
(Tocomp and Objectnotpermitted)	INTRANSTOCOMP
Np-tocomp	NPTOCOMP
Bareinfinitivecomp	BARE-INF-COMP
Makecomp	MAKECOMP
Participlecomp	PARTICIPLECOMP
Pronouns	
Nopostmodifiers	NOPOSTMODIFIERS
(Pronoun and Location)	NONPOBJECT
Nouns	
Determinerrequired	DETREQUIRED
(Noun and Uncountable)	NONCOUNT
(Noun and Thatcomp)	FACTIVE
Other things	
Limiter	PREDETADV
(Preposition and Objectnotrequired)	BAREPREP
Ppobject	PPOBJECT

11.1. Immovableparticles and Particles

For these properties, part of the RUS version of the property value is redundant. Therefore, it will improve the legibility of the ML if we manufacture these parts of the values by rule. This generally involves gluing the word onto the property value. As previously, W stands for the word spelling.

(Immovableparticles P1 ... Pn)	IMMOVABLEPARTICLES ((W W\P1)... (W W\Pn))
(Particles P1 ... Pn)	PARTICLES ((W W\P1) ... (W W\Pn))

11.2. Compounds and Multiples

These two RUS properties carry redundant lexicon-internal information; thus they can be supplied entirely by the translation rules. The rules are as follows:

If there exists an entry W1% W2, and there is no entry W2, add (W2 (W1\W2)) to the entry for W1 as a value of the property MULTIPLES. The output should look like: MULTIPLES ((W21 (W1\W21)) ... (W2n (W1\W2n)))

If there exists an entry W1% W2, and there is an entry W2, add (W2 W1\W2) to the entry for W1 as a value of the property COMPOUNDS. The output should look like: COMPOUNDS ((W21 W1\W21) ... (W2n (W1\W2n)))

References

[Bates 83] Bates, M. and R. J. Bobrow.
The RUS Parser User's Guide.
Technical Report, Bolt, Beranek and Newman, Inc., April, 1983.

[Bates 84a] Bates, M. and R. Ingria.
Dictionary Documentation.
Technical Report, Bolt Beranek and Newman, Inc., September, 1984.

[Bates 84b] Bates, M. and R. Ingria.
Dictionary Package Documentation.
Technical Report, Bolt Beranek and Newman, Inc., September, 1984.

[Cumming 86] Cumming, Susanna and Robert Albano.
A Guide to Lexical Acquisition in the JANUS system.
Technical Report, USC/ISI, January, 1986.

[Ingria 85] Ingria, Robert.
Dictionary Morphology Documentation.
Technical Report, Bolt, Beranek and Newman, 1985.
In Progress.

END

DTIC

4-86